



UNIVERSITY OF
GEORGIA

GlyGen

Computational and Informatics Resources for Glycoscience

Rupali Mahadik, Rene Ranzinger, Nathan Edwards, Robel Kashay, Raja Muzumder, William York.

Complex Carbohydrate Research Center, University of Georgia, Athens, GA, USA.

The George Washington University, Washington, DC, USA.

THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

ABSTRACT

- Advancing our understanding of the roles that glycosylation plays in development and disease is hindered by the diversity of the data that must be integrated to gain insight into these complex phenomena.
- GlyGen retrieves information from multiple international data sources and integrates and harmonizes this data into a single database
- This web portal allows exploring this data and performing unique searches that cannot be executed in any of the integrated databases alone.

DATA / API

- Data from major genomics, proteomics resources data generators is collected and downloaded.
- All information is integrated into our data model. As part of this integration workflow, datasets are translated into consistent name spaces and have to pass through intensive quality control workflows.
- Direct access to GlyGen data is possible from the Data page, web service API and SPARQL endpoint.

Poster B113

COLLABORATIONS

Kiyoko Aoki-Kinoshita PubChem Evan Bolton
 Maria Martin UniCarbKB Matthew Campbell

WHY - GLYGEN ?

- Up-to-date glycobiology-related information from diverse data sources.
- Intuitive web portal to browse and search for knowledge in glycobiology.
- Protein and glycan data for Human, Rat, Mouse

WHAT - NEXT ?

- Data from other species and other resources.
- Improved help system for users.
- 3D structures for glycans, proteins and glycoproteins.
- improved widgets to navigate and display data

ACKNOWLEDGEMENT

- The GlyGen project is supported by an NIH Common Fund grant (U01 GM125267-01)
- The project is an international collaboration with many people involved: Radoslav Goldman, Judith Blake, Michael Pierce, Karen Ross, Rob Woods, Ten Feizi, Jeff Gildersleeve, Richard Cummings and many people in their groups.

WEB - PORTAL

<https://www.glygen.org/>



The GlyGen portal allows browser-based access to multidisciplinary data integrated from diverse sources. The portal is divided into three entry points: glycans, proteins and glycoproteins. For each type of molecule an extensive list of information integrated from the different data sources is displayed. Protein and glycan information pages are interlinked with each other allowing the user to navigate from a protein to its glycans and vice versa. Extensive search interfaces have been implemented for each type of molecule allowing the user to search by molecule IDs, common namespaces, chemical properties or relationships to other molecules. In preparation for the project, an extensive user survey was performed, leading to ~250 use cases from more than 50 international scientists. A user-friendly interface called "Quick Search" has been implemented to address some of the complex questions that require data and querying across multiple domains.

GLYCOPROTEIN

- Simple or advanced search for Glycan, Protein, Glycoprotein
- Quick Search to answer user-specific question using predefined queries.
- Each search generates a set of search results - glycan or (glyco)protein list
- Searches can be bookmarked, reviewed and modified.
- Download option to save and process search results.
- Details pages provide an exhaustive overview about all the information about a glycan or (glyco)protein
- For each glycan the system provides: species annotation, a list of common motives, sequence formats, publications and proteins that have this glycan attached
- For each protein the system provides: extensive data, including species, function, sequence, cross references, homologs, isoforms, mutations, and expression levels
- All information in the webpages is tagged and linked with the original data sources, allowing users to find the source of each piece of information and link out to additional data that has not been integrated into GlyGen
- Protein and glycan information pages are interlinked with each other allowing the user to navigate from a protein to its glycan's and vice versa